

Probabilistic wind power forecasting models

Soumaya Elkantassi, Evangelia Kalligiannaki,

Raúl Tempone (Director, Center for UQ, KAUST)

Alexander von Humboldt Professorship Awardee (RWTH, Germany)

<http://sri-uq.kaust.edu.sa>, http://stochastic_numerics.kaust.edu.sa

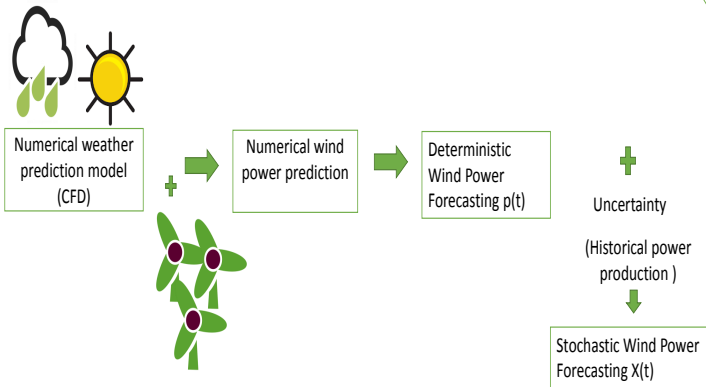
King Abdullah University of Science & Technology, Saudi Arabia



- 1 About KAUST
- 2 Numerical Weather Prediction & historical wind power production
- 3 Modelling and Statistical Inference
- 4 Application to aggregate wind power production in Uruguay
- 5 Forecasting Solar power production

Probabilistic wind power forecasting

Stochastic Wind Power Forecasting



- Provide stochastic forecast analogue to $p(t)$ \Rightarrow scenarios of wind power forecasting with confidence bands.
- Stochastic optimization problems: management of electricity costs. AUST

NWP and historical wind power production

Numerical predictions are provided by **three different predictions** in Uruguay during the following periods.

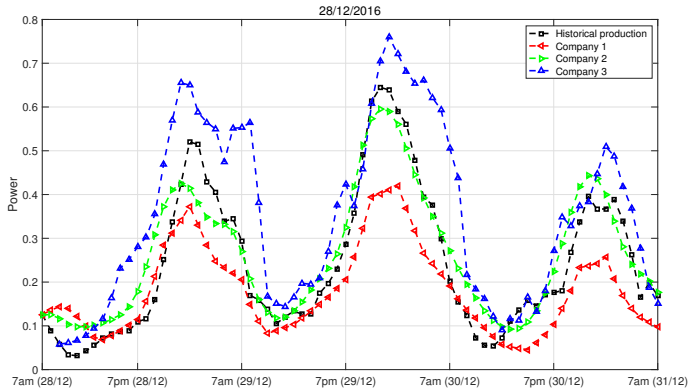
Company	N	From	To
1	324	04/04/2016	09/03/2017
2	417	08/01/2016	09/03/2017
3	248	01/01/2016	28/12/2017

Table: Available daily predictions

The data sets represent the value of aggregate wind power (in MW), normalized with the nominal capacity.

Normalized historical power d_{ji} and numerical forecast p_{ji}^k

$i = 1, \dots, T$, $T = 1, \dots, 72$ the time horizon, $j = 1, \dots, N$, N the number of days available, and three different NWFs $k = 1, 2, 3$.



Modelling

Let, $X(t)$, denote the stochastic forecast analogue to $p(t)$.

Assume it is the solution of the parametrized SDE

$$\begin{cases} dX(t) = b(X(t); \theta)dt + \sigma(X(t); \theta)dW(t), & t > 0, \\ X(0) = x_0 \end{cases}$$

$X(t) \in [0, 1]$ and θ denotes a vector of parameters.

- **Form of the drift:** $X(t)$ is 'close' to $p(t)$. In expectation, $X(t)$ tends to $p(t)$

$$b(X(t), t; \theta) = -\theta(t)(X(t) - p(t))$$

$\theta(t)$: the rate by which the variable reverts to $p(t)$ in time.

- **Form of the diffusion:** the diffusion vanishes at the boundaries

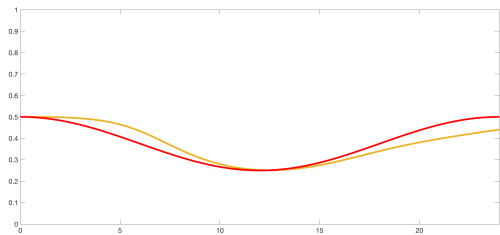
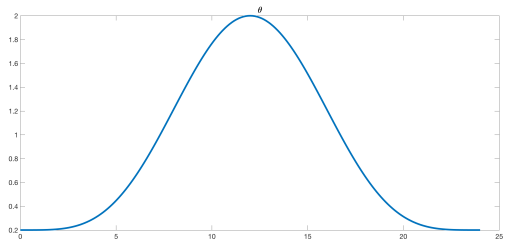
$$\sigma(X(t); \theta) = \sqrt{2\alpha(t)\theta(t)p(t)(1-p(t))X(t)(1-X(t))}$$

$\alpha(t)$: controls the path variability.

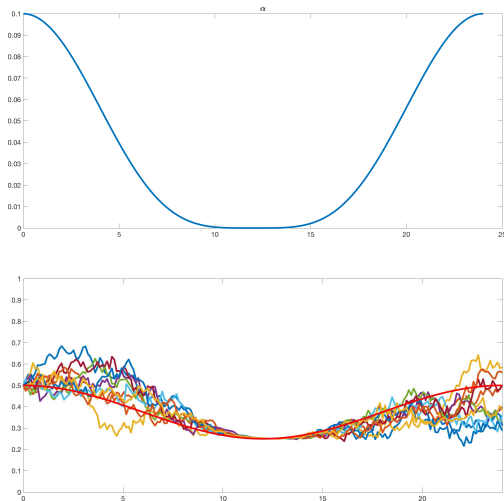
Some observations:

- emphasis in not only band but also reasonable paths (for instance time correlation) that represents the wind power ensemble
- ensemble not centered around the prediction value
- flexibility provided by time dependence of α and θ .

Deterministic case:



Stochastic case:



Drawback: Ensemble may have centering issues wrt to prediction.

- **Improved form of the drift:**

Goal: better centering of the ensemble around the prediction value

Approach: $X(t)$ is 'close' to $p(t)$ AND also tracks its derivative.

We require that

$$b(p(t), t; \theta) = \frac{dp}{dt}$$

$$b(X(t), t; \theta) = \frac{dp}{dt} - \theta(t) \left(X(t) - p(t) \right)$$

$\theta(t)$: the rate by which the variable reverts to $p(t)$ in time.

Stochastic case, tracking prediction derivative:

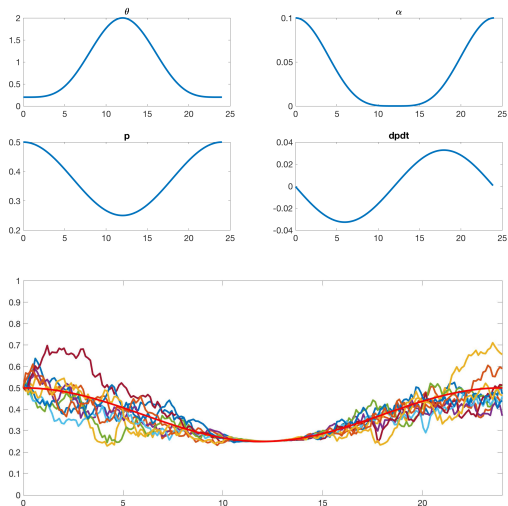


Figure: Ensemble exhibits better centering wrt to prediction.

More observations:

- Upper $U(t)$ and Lower $L(t)$ bounds (smooth bands), satisfying for all times $L(t) < p(t) < U(t)$, may be imposed provided that
 - the diffusion σ vanishes at $X(t) = U(t)$ and $X(t) = L(t)$.
For instance we may take

$$\sigma_t^2(x) \propto (U(t) - x)(x - L(t))$$

- the drift b satisfies

$$b_t = \frac{dp}{dt}, \text{ at } X(t) = p(t)$$

$$b_t < dU/dt, \text{ at } X(t) = U(t)$$

$$dL/dt < b_t, \text{ at } X(t) = L(t)$$

Example [Symmetric, relative band]

Here

$$L(t) = p(t)(1 - R(t)),$$

$$U(t) = p(t)(1 + R(t)),$$

and we may take, for $\theta(t) > 0$,

$$b_t = \frac{dp}{dt} - \theta(t) \tan \left(\frac{\pi}{2R(t)} \left(\frac{x}{p(t)} - 1 \right) \right)$$

Statistical Inference for SDE's (Ongoing Work)

$$\mathcal{D} = \{X_{t_i}\}_{i=0}^N$$

Maximum Likelihood Estimation

– Exact:

$$\mathcal{L}(\theta; \mathcal{D}) = Q^\theta(X_{t_0}, \dots, X_{t_n}) = \prod_{i=1}^n \rho(t_i, X_{t_i} | t_{i-1}, X_{t_{i-1}}; \theta) \rho(X_{t_0})$$

$\rho(t, x | s, y; \theta)$ is the solution of the Fokker-Planck equation

$$\frac{\partial}{\partial t} \rho(t, x | s, y; \theta) = -\frac{\partial}{\partial x} (b(x; \theta) \rho(t, x | s, y; \theta)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\sigma^2(x; \theta) \rho(t, x | s, y; \theta))$$

– Approximate: $L(\theta; \mathcal{D})$

- the process $\{X_t\}$ by a 'simpler' process, e.g. by a Gaussian process,
- the process $\{X_t\}$ by a discrete scheme, e.g. Euler-Maruyama, Milstein, Ozaki,
- the transition probability, e.g. simulated likelihood, Hermite polynomial expansion

Observation equation:

$$\mathbf{D}^{(j)} = \mathbf{G}^{(j)} + \epsilon^{(j)},$$

- $\mathbf{G}^{(j)} = (G^{(j)}(t_1), \dots, G^{(j)}(t_N))$

$$\mathbf{d}_{jN} = \left\{ \mathbf{d}^{(j)} = \{d_{ji}\}_{i=1}^N \right\}_{j=1}^J$$

$G^{(j)}(t)$ Gaussian approximation of $X^{(j)}(t)$ (solution of the SDE with $p(t) = p^{(j)}(t)$), defined by:

- $\mu^{(j)}(t) = \mathbb{E}[X^{(j)}(t)],$
- $v^{(j)}(t) = \mathbb{E}[(X^{(j)}(t) - \mu^{(j)}(t))^2]$
- $v^{(j)}(t, s) = \mathbb{E}[(X^{(j)}(t) - \mu^{(j)}(t))(X^{(j)}(s) - \mu^{(j)}(s))]$
- $\epsilon^{(j)} \sim \mathcal{N}(0, \Sigma^{\epsilon, (j)})$, denotes the measurement error

Approximate likelihood

The approximate likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{d}_{JN}) = \prod_{j=1}^J (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta})|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))^T [\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta})]^{-1} (\mathbf{d}^{(j)} - \boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}))}$$

- $\boldsymbol{\mu}^{(j)}(\boldsymbol{\theta}) = (\mu^{(j)}(t_1), \dots, \mu^{(j)}(t_N))$,
- $\boldsymbol{\Sigma}^{(j)}(\boldsymbol{\theta}) = \mathbf{V}^{(j)}(\boldsymbol{\theta}) + \boldsymbol{\Sigma}^{\epsilon, (j)}(\boldsymbol{\theta})$,
 - $\mathbf{V}^{(j)}(\boldsymbol{\theta})$ is the covariance matrix,
 $[\mathbf{V}^{(j)}(\boldsymbol{\theta})]_{kl} = v^{(j)}(t_k, t_l)$, $k, l = 1, \dots, N$,
 - $\boldsymbol{\Sigma}^{\epsilon, (j)}(\boldsymbol{\theta}) = \phi^2 \mathbf{I}$

$|\boldsymbol{\Sigma}|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}$.

$\boldsymbol{\theta} = (\theta(\cdot), \alpha, \phi)$ the vector of parameters

Optimization problem

$$\boldsymbol{\theta}^* = \operatorname{argmin}\{-\log(L(\boldsymbol{\theta}; \mathbf{d}_{JN}))\}$$

Fitting of time independent parameters

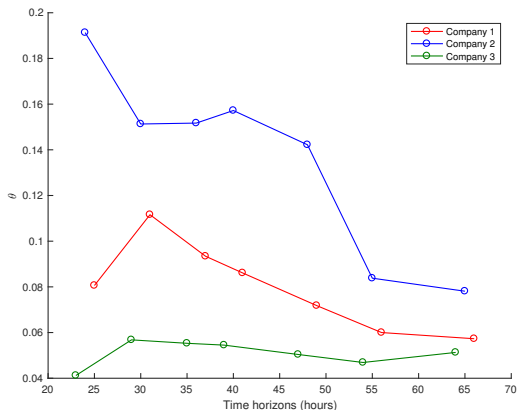


Figure: Comparison between the estimated constant mean reversion rate (θ) for the three companies

Optimal combination of forecasts

Linear combination of the Companies 2 and 3 point forecasts

$$p_\lambda(t) = \lambda p^{(2)}(t) + (1 - \lambda)p^{(3)}(t), \lambda \in [0, 1]$$

Find the optimal $(\theta, \alpha, \phi, \lambda)$ for which $X_\lambda(t)$, best approximates the historical data \mathcal{D} .

$$dX_\lambda(t) = -\theta(X_\lambda(t) - p_\lambda(t))dt + \sigma(X_\lambda(t); \theta, \alpha, \lambda)dW(t)$$

Optimal combination of forecasts

Linear combination of the Companies 2 and 3 point forecasts

$$p_\lambda(t) = \lambda p^{(2)}(t) + (1 - \lambda) p^{(3)}(t), \lambda \in [0, 1]$$

Find the optimal $(\theta, \alpha, \phi, \lambda)$ for which $X_\lambda(t)$, best approximates the historical data \mathcal{D} .

$$dX_\lambda(t) = -\theta(X_\lambda(t) - p_\lambda(t))dt + \sigma(X_\lambda(t); \theta, \alpha, \lambda)dW(t)$$

MLE estimates for the constant rate SDE model

Horizon T	$(\theta, \alpha, \lambda)$	log likelihood
24	(0.167, 0.352, 0.910)	1.37e+04
36	(0.158, 0.396, 0.900)	1.01e+04
48	(0.140, 0.409, 0.914)	1.35e+04

Optimal λ is approximately 0.9, which indicates that **company 2 provides a better quality**.

Fitting of time varying rate for Company 2

Recall the model SDE

$$dX(t) = -\theta(t) (X(t) - p(t)) dt + \sigma(X(t); \theta(t), \alpha) dW(t)$$

- Expected properties of the rate function
 - Decay with increasing time -
the predictability of NWP $p(t)$ decays with time
 - Positive
- Proposed models with parameterizations of the rate $\theta(t)$

SDE	Rate function	Number of parameters
Constant	$\theta_0(t) = \theta_0$	3
Linear	$\theta_1(t) = -\theta_0 t + \theta_1$	4
Exponential	$\theta_2(t) = \theta_0 \exp^{-\theta_1 t}$	4
Rational	$\theta_3(t) = \frac{\theta_0}{(\theta_1 + t)}$	4
Power	$\theta_4(t) = \theta_0 t^{-\theta_1}$	4

Time varying rate models for company 2

Time-horizon $T = 24$ hours.

Initial values for (θ_0, θ_1) are chosen by fitting the chosen function to the constant rate MLE for different time horizons and $(\alpha^0, \phi^0) = (0.2, 0.1)$.

SDE	Initial values (θ_0^0, θ_1^0)	MLE estimates $(\theta_0, \theta_1, \alpha, \phi)$	CPU time(seconds)
Constant	(0.400, --)	(0.191, --, 0.277, 0.007)	2850
Linear	(0.003, 0.250)	(0.003, 0.226, 0.283, 0.007)	3148
Exponential	(0.303, 0.020)	(0.210, 0.010, 0.284, 0.007)	3840
Rational	(7.377, 14.35)	(21.06, 125.4, 0.418, 0.004)	5864
Power	(2.001, 0.733)	(5.607, 0.797, 0.192, 0.096)	6754

Remarks

- Constant, Linear and Exponential rate models give 'the same' α and ϕ values
- Best in computational cost is the constant rate model

Model selection

Model selection using Information Criteria

- Akaike Information Criterion - AIC (the lower AIC the better)
- Bayesian Information Criterion - BIC (the lower BIC the better) and the
- log-likelihood $\log(L)$ (the larger $\log(L)$ the better)

SDE	$\log(L)$	AIC	BIC
Constant	24128	-48250	-48238
Linear	24139	-48270	-48254
Exponential	24135	-48262	-48246
Rational	24044	-48080	-48064
Power	18469	-36930	-36914

The Linear rate model, $\theta_1(t) = -\theta_0 t + \theta_1$, is the preferred model.

The Exponential rate model, $\theta_2(t) = \theta_0 e^{-\theta_1 t}$ is close.

Model Validation

We fit the model with a training set to get:

	$\theta_0, \theta_1, \alpha, \phi$
Linear	(0.002, 0.190, 0.302, $8.5e - 3$)
Exponential	(0.174, 0.007, 0.303, $8.5e - 3$)

We solve one-dimensional Fokker-Planck equation for a randomly chosen path within the testing set

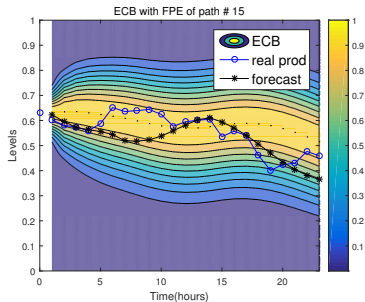
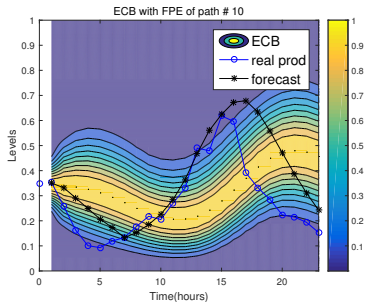



Figure: Validation of the numerical results with the linear model 

Validation on the total testing set

We measure the **number of exceedances** of the 95% Confidence Interval.

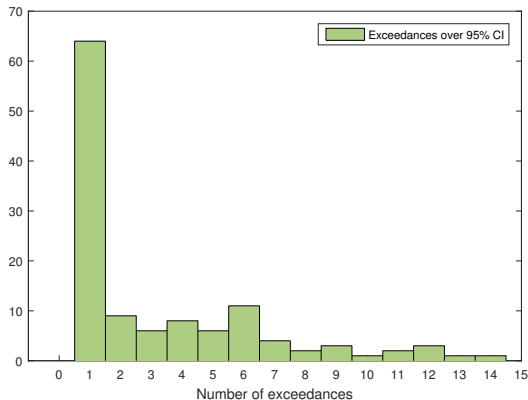


Figure: Validation of the numerical results with the exponential model

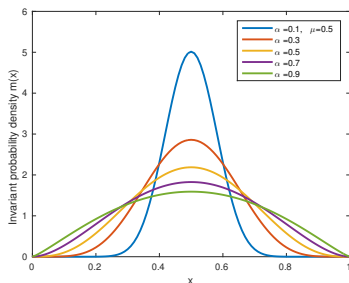
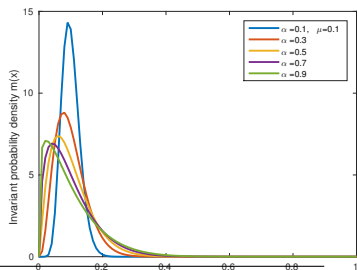
Case study: $p(t) = \mu = \text{constant}$, Invariant distribution

$$\begin{cases} dX(t) = -\theta(X(t) - \mu)dt + \sqrt{2\theta\alpha\mu(1-\mu)X(t)(1-X(t))}dW(t), & t > 0 \\ X(0) = x_0 \end{cases}$$

- Mean reverting $d\mathbb{E}[X(t)] = -\theta(\mathbb{E}[X(t)] - \mu)dt$
- Autocorrelation function of the process $\text{cor}(X(t), X(t + \tau)) = e^{-\theta\tau}$

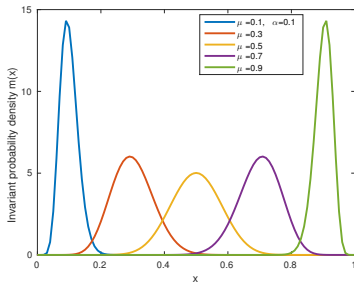
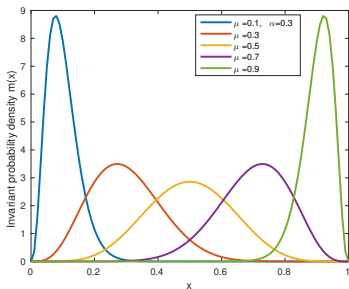
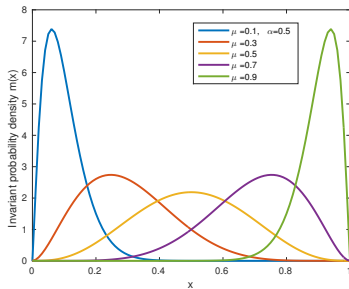
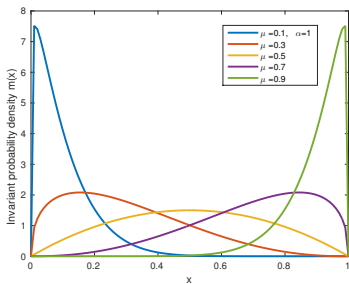
Invariant probability density $m(x)$ exists¹ in $x \in (0, 1)$, and

- is a Beta distribution $\mathcal{B}(\frac{1}{\alpha(1-\mu)}, \frac{1}{\alpha\mu})$. (Independent of θ)
- with mean μ and variance is $\frac{1}{\alpha} \frac{((1-\mu)\mu)^2}{1+\alpha\mu(1-\mu)}$



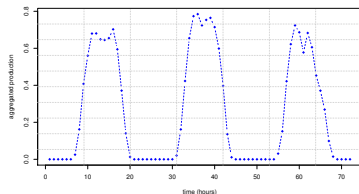
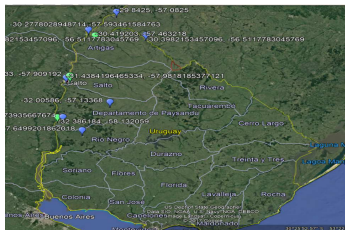
¹Forman, et.al. Scand. Jour. Stat., 2008

Invariant distribution. Case study: $p(t) = \mu = \text{constant}$



Probabilistic forecast of Solar power production

- Predict the solar power production given the influence of clouds, aerosols, and other atmospheric constituents.



- We will consider models based on **Stochastic Differential Equations (SDEs)**.
- We consider models based on physical features of interest:
 - 1 Model 1 Clear Sky model.
 - 2 Model 2: A model involving cloudiness predictions.

Description of the data

- We have available three data sources:
 - (1) The hourly solar power production of three photovoltaic farms on the 366 days of the year 2016, located in three departments of Uruguay. These data are normalized by the installed power (MW) of each farm.

Farm	Department	Installed capacity	Operativity
Alto Cielo	Artigas	20	From 1st March, 2016
La Jacinta	Salto	50	All 2016
Raditon	Paysandú	8	All 2016

- (2) The hourly cloudiness during the year 2016, provided by Instituto Uruguayo de Meteorología (INUMET). These data are available for seven meteorological stations enclosing the three farms.
 - (3) Daily sunset and sunrise times during 2016.
- The aggregated production is calculated as the sum of the three productions in MW and divided by the total installed capacity (78 MW, after March 1st).

Model 1: Clear sky

Then, for the SDE

$$dY(t) = \beta \frac{\varepsilon_t}{dt} dt - \theta(Y_t - \beta \varepsilon_t) dt + \sigma(Y_t, \theta) dW_t,$$

- The function ε_t is an upper envelope of the observed process, fitted from the observed data.
- ε_t represents the “Clear sky” production of the plant, i.e. the solar production with absence of cloudiness.

-

$$\sigma(y; \theta) = \sqrt{2\theta\alpha y(\varepsilon_t - y)}.$$

The upper envelope described

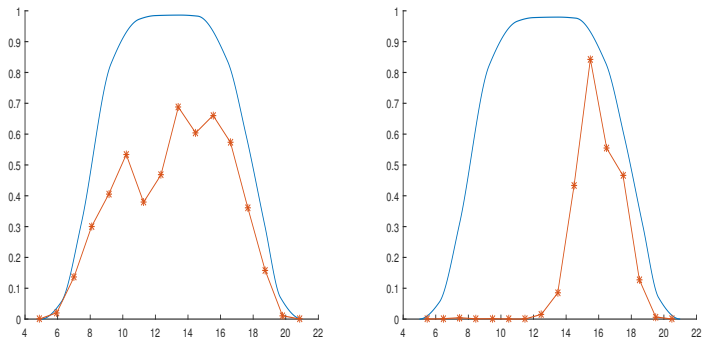
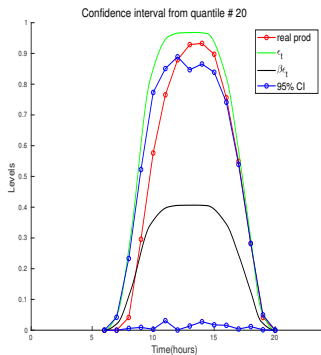
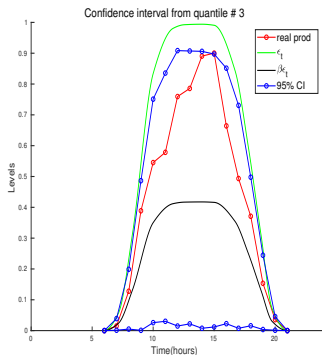


Figure: Upper envelope (blue), observed data (orange).

Maximum likelihood estimation and validation

- From generated samples we also compute the 0.025 and the 0.975 quantiles for each $t = t_0^{(j)}, \dots, t_f^{(j)}$. **No derivative tracking though ...**



95% Confidence Intervals with quantiles

Model 2: Cloudiness states

Then, for the SDE

$$dY(t) = \beta(t) \frac{\varepsilon_t}{dt} dt - \theta(Y_t - \beta(t)\varepsilon_t)dt + \sigma(Y_t, \theta)dW_t,$$

Markovian switching model

- We count with hourly state sky data for three locations, each one near to each Photovoltaic plant.
- With this data we will fit a jump process R_t with state space $\mathcal{S} = \{0, 1, 2, 3, 4\}$ given by

\mathcal{S}	Cloudiness
0	Clear sky
1	Half cloudy
2	Cloudy
3	Completely cloudy
4	Obstructed from view

- We shall fit the jump process R_t transition probabilities from the historical cloudiness data.
- Our aim is to find the optimal values $\beta(t) \in \{\beta_0, \dots, \beta_1\}$

Summary: modeling renewable sources with SDEs

Wind power (ongoing work)

- Aggregate wind power forecasting quantifying uncertainty of NWP
- Founded on Statistical first principles (max. Lik., Bayesian)
- Produces power production path scenarios
- Flexible approach can
 - accommodate day/night effects
 - introduce bands for variability a.s.
- Indirect inference using approximate Likelihood to accelerate computations
- If local forecasts and historical values are available, then we can distinguish among different farms (systems of SDEs) taking into account their correlation.

Summary: modeling renewable sources with SDEs

Solar power (ongoing work)

- Without NWP involved the approach is limited, better to use couples of predictions and historical values! **Can anyone provide these data?**
- Based on clear sky model, which we fit form historical observations
- Effect of cloudiness, that is highly volatile but can be mitigated by using meteorological forecasts.

Thank you!